Bases de Datos Avanzadas

M. en C. Alejandro Botello Castillo

**Práctica no. 8:**

Diseño de una Base de Datos Multidimensional

**Número de Práctica: 1**

**Nombre de la Práctica: Diseño de una Base de Datos Multidimensional**

REQUISITOS PREVIOS

Para la realización de la práctica se deberá tener instalado en la computadora el software Oracle XE y [Data Integration](https://drive.google.com/file/d/1g3kESPagO1HId4I_yAKYU291-M4w2seg/view?usp=sharing). El software será proporcionado por el profesor. Debe seguir los siguientes pasos para la instalación y puesta en marcha:

1. Descomprimir el archivo [*OracleXE11g.zip*](https://drive.google.com/file/d/1P9EVa3eU5Fp3v80ojSSzDFpSzoH6nMaH/view?usp=sharing) en la computadora local. De la carpeta llamada *DISK1* ejecutar el archivo *setup.exe* y permitir la ejecución del instalador. Interfaz de usuario gráfica, Aplicación

   Descripción generada automáticamente
2. De las pantallas que aparecen, aceptar el contrato de licencia para poder instalar el software. Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

   Descripción generada automáticamente
3. Indicar la carpeta en donde se hará la instalación (se recomienda dejar por omisión). Verificar que se cumple con el espacio en disco necesario (se requiere poco más de 500 MB para el software y la base de datos).

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. **IMPORTANTE**. Es necesario establecer la contraseña del usuario administrador del sistema (*system*).Considere GUARDARLA EN LUGAR SEGURO. Interfaz de usuario gráfica, Texto

   Descripción generada automáticamente
2. Guardar los parámetros de configuración porque son necesarios para la conexión en prácticas posteriores (el profesor indicará los momentos en que sean necesarias), en particular los puertos en que están ejecutando los procesos.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. Terminada la instalación, se deberá verificar que los servicios de bases de datos se están ejecutando correctamente. Abrir el *Administrador de Tareas* de Windows y ver la pestaña de **servicios**. Comprobar que los servicios *OracleServiceXE* y *OracleXETNSListener* estén ejecutando correctamente. Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

   Descripción generada automáticamente
2. Descomprima el archivo descargado *data-integration.zip* en alguna carpeta del sistema de archivos local. Abra una consola de línea de comandos de Windows y vaya al directorio *data-integration* recién descomprimido (con el comando cd). Configure las variables de ambiente JAVA\_HOME, CLASSPATH y PATH para la ruta en que se encuentre la MV de Java (el software ya incluye la máquina virtual versión 7, por lo que no es necesario descargarla o tenerla instalada).

set JAVA\_HOME=[drive]:[rutalocal]\data-integration\jdk\jre

set CLASSPATH=[drive]:[rutalocal]\data-integration\jdk\jre\lib

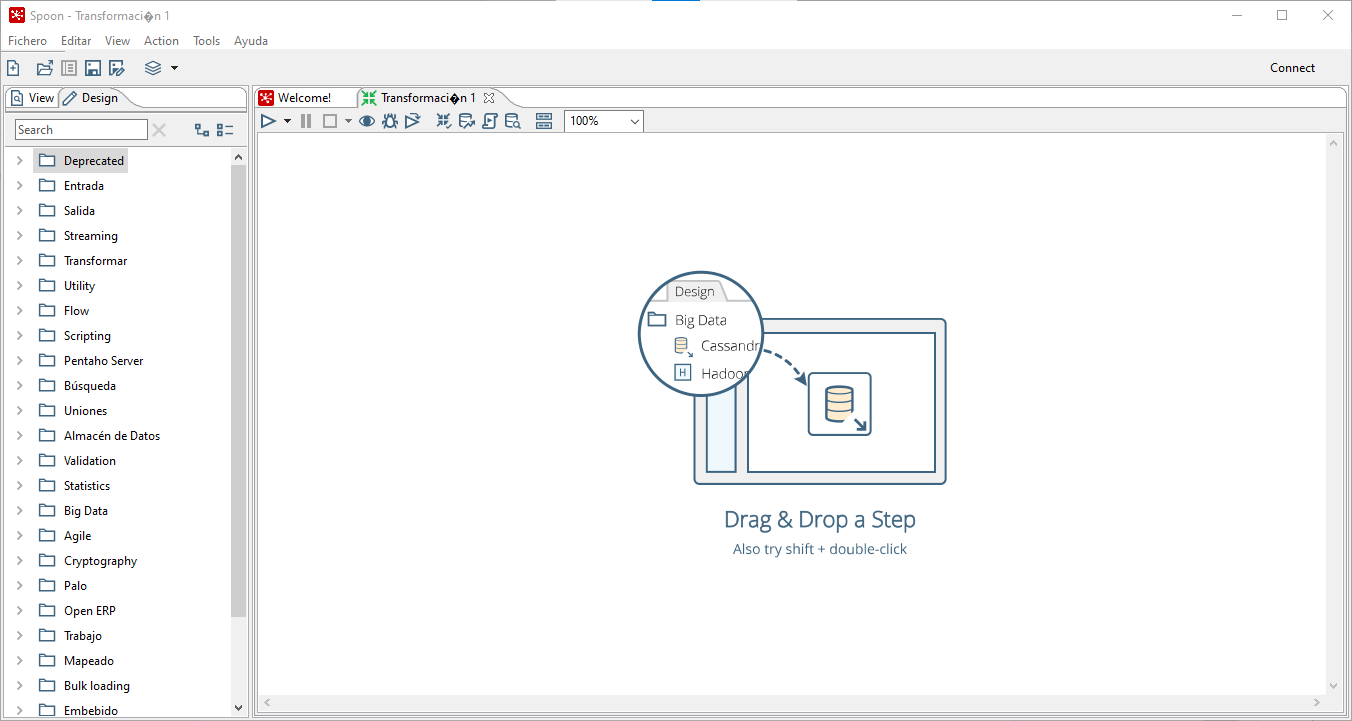
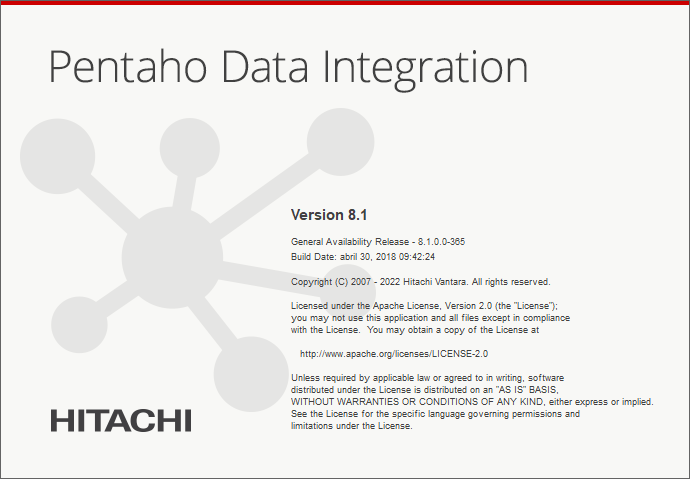
set PATH=[drive]:[rutalocal]\data-integration\jdk\jre\bin

y ejecute el archivo *set-pentaho-env.bat* para configurar el programa.

set-pentaho-env.bat

De la siguiente ruta [drive]:\oraclexe\app\oracle\product\11.2.0\server\jdbc\lib, localice el archivo ojdbc6.jar y cópielo a la carpeta [drive]:[rutalocal]\data-integration\lib

Una vez concluida la configuración, ejecute el programa, escribiendo en la terminal el comando Spoon.bat. Espere un poco hasta que la aplicación cargue por completo.



1. Descargue el archivo [*ventas.mdb*](https://drive.google.com/file/d/1U2TqKs5k60ZHq_RH81wmp7xYCfWoIK6C/view?usp=sharing) que representa una base de datos operacional de Access, y guárdelo en una carpeta accesible en la computadora local.

MARCO TEÓRICO

De acuerdo a lo visto en clase, se emplearán los conceptos del modelado multidimensional aplicado a un ejemplo de una base de datos operacional, y la inserción de datos empleando una herramienta ETL.

DESARROLLO DE LA PRÁCTICA:

Suponga el siguiente modelo de una base de datos de una empresa, que representa el registro de las ventas de un negocio para determinados clientes en los productos que comercializa. El archivo que contiene está base de datos es *ventas.mbd* y es de la aplicación Access y acompaña al software proporcionado por el profesor.

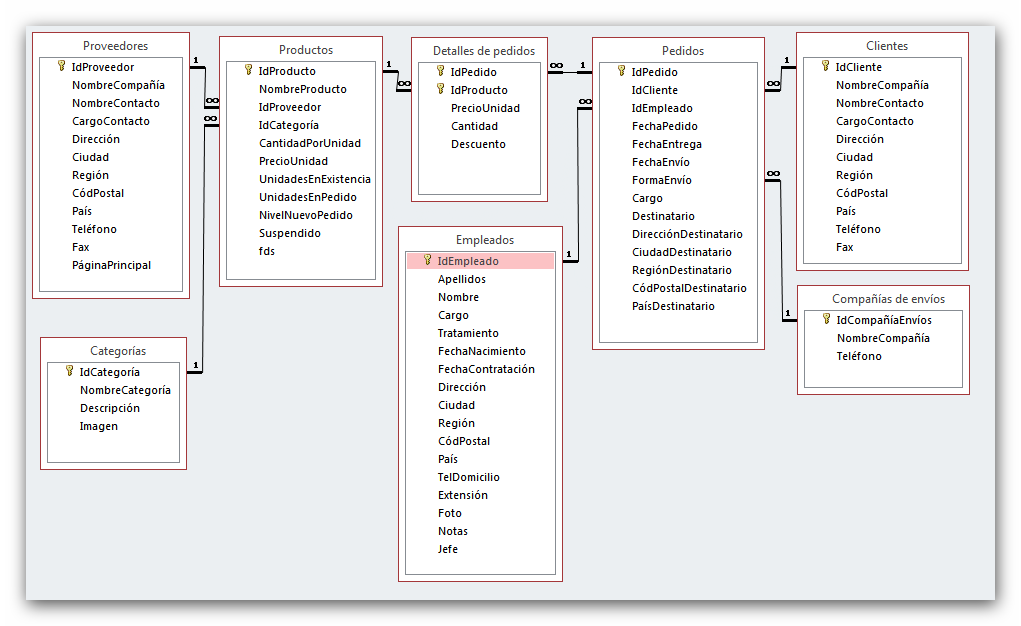


Figura 1. Modelo relacional de la base de datos operacional

Se observa que la base de datos está diseñada bajo el modelo relacional y que cumple con la normalización.

La empresa está interesada en hacer análisis histórico de sus datos generados, por lo que se ha decidido almacenar estos datos en un modelo multidimensional, que resuelva las siguientes consultas:

1. Se desea saber el total de productos vendidos, por categoría y por producto.
2. El total de ventas de los productos vendidos, por categoría y por producto.
3. El total de ventas de los productos, por categoría y por tipo (estado) de cliente.
4. Los cinco productos más vendidos por categoría.
5. Los cinco clientes más compradores por ciudad.
6. Las cinco fechas en donde se obtuvieron más ventas por cada año.
7. Las cinco fechas en donde se venden más productos cada año.
8. Los totales de las ventas de productos por categorías por cada mes del año.

De acuerdo con las preguntas interesantes del negocio, es necesario modelar hacia multidimensional la base de datos anterior. El modelo propuesto es el siguiente:

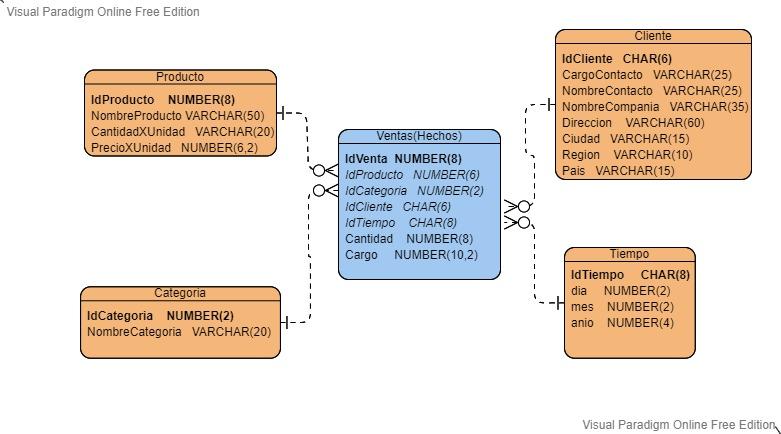


Figura 2. Modelo multidimensional a implementar

9.- Mediante el software Oracle XE (ya debe estar instalado en la computadora de trabajo), crear el modelo multidimensional anterior. Inicie el servidor de Oracle con el icono *Start Database* del grupo de programas *Oracle Database XE 11g* en el menú de inicio de Windows. Espere a que inicie el sistema y mantenga la ventana de comandos abierta pero minimizada.

Captura de pantalla de un celular

Descripción generada automáticamente

2.- Abra una ventana de comandos *Run SQL Command Line* y ejecute las siguientes sentencias:

SQL> connect / as sysdba;

10.- Por cuestiones de seguridad, agregue un nuevo usuario para el servidor de bases de datos, con nombre **dwh**y contraseña **dwh**. Escriba las siguientes instrucciones:

SQL> create user dwh identified by dwh;

SQL> grant connect, resource to dwh;

11.- Mantenga abierta esa ventana y abra una nueva ventana terminal de línea de comandos. Escriba las siguientes sentencias:

SQL> connect dwh/dwh;

CREATE TABLE CATEGORIA

(IDCATEGORIA NUMBER(2,0) NOT NULL,

NOMBRECATEGORIA VARCHAR2(50),

CONSTRAINT CATEGORIA\_PK PRIMARY KEY (IDCATEGORIA));

CREATE TABLE CLIENTE

(IDCLIENTE VARCHAR2(6) NOT NULL,

NOMBRECONTACTO VARCHAR2(25),

CARGOCONTACTO VARCHAR2(40),

NOMBRECPIA VARCHAR2(40),

DIRECCION VARCHAR2(60),

CIUDAD VARCHAR2(15),

REGION VARCHAR2(40),

PAIS VARCHAR2(15),

CONSTRAINT CLIENTE\_PK PRIMARY KEY (IDCLIENTE));

CREATE TABLE PRODUCTO

(IDPRODUCTO NUMBER(8,0) NOT NULL,

NOMBREPRODUCTO VARCHAR2(50),

CANTIDADXUNIDAD VARCHAR2(50),

PRECIOXUNIDAD NUMBER(10,2),

CONSTRAINT PRODUCTO\_PK PRIMARY KEY (IDPRODUCTO));

CREATE TABLE TIEMPO

(IDTIEMPO VARCHAR2(8) NOT NULL,

ANIO NUMBER(4,0),

MES NUMBER(2,0),

DIA NUMBER(2,0),

CONSTRAINT TIEMPO\_PK PRIMARY KEY (IDTIEMPO));

CREATE TABLE VENTAS

(IDVENTA NUMBER(8,0) NOT NULL,

CANTIDAD NUMBER(5,0),

CARGO NUMBER(10,2),

PRODUCTO\_IDPRODUCTO NUMBER(8,0),

CATEGORIA\_IDCATEGORIA NUMBER(2,0),

CLIENTE\_IDCLIENTE VARCHAR2(6),

TIEMPO\_IDTIEMPO VARCHAR2(8));

ALTER TABLE VENTAS ADD CONSTRAINT VENTAS\_CATEGORIA\_FK FOREIGN KEY (CATEGORIA\_IDCATEGORIA) REFERENCES CATEGORIA (IDCATEGORIA) DISABLE;

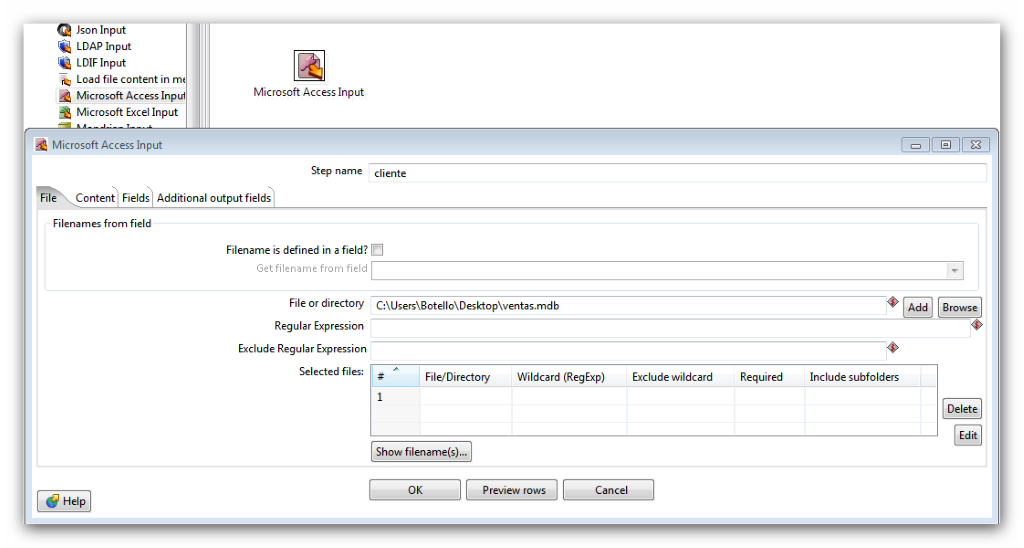
ALTER TABLE VENTAS ADD CONSTRAINT VENTAS\_CLIENTE\_FK FOREIGN KEY (CLIENTE\_IDCLIENTE) REFERENCES CLIENTE (IDCLIENTE) DISABLE;

ALTER TABLE VENTAS ADD CONSTRAINT VENTAS\_PRODUCTO\_FK FOREIGN KEY (PRODUCTO\_IDPRODUCTO) REFERENCES PRODUCTO (IDPRODUCTO) DISABLE;

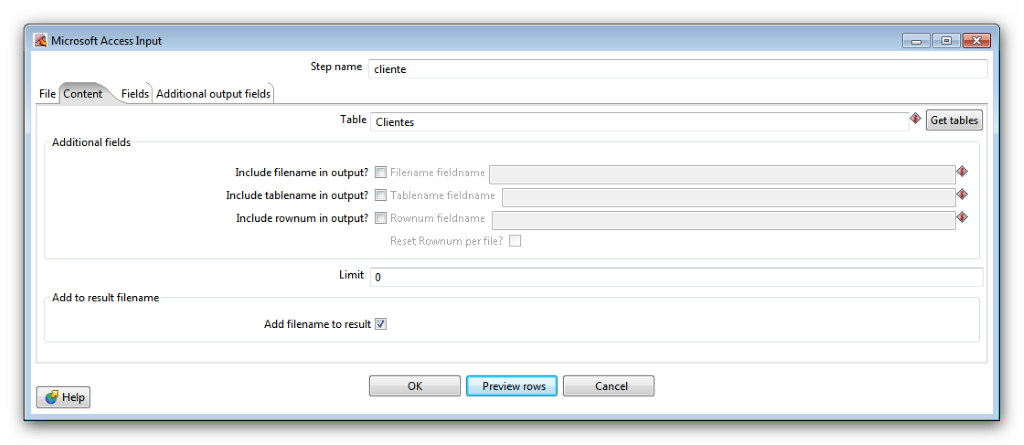
ALTER TABLE "VENTAS" ADD CONSTRAINT VENTAS\_TIEMPO\_FK FOREIGN KEY (TIEMPO\_IDTIEMPO) REFERENCES TIEMPO (IDTIEMPO) DISABLE;

12.- Cambie a la aplicación Pentaho Data Integration y realice los siguientes pasos:

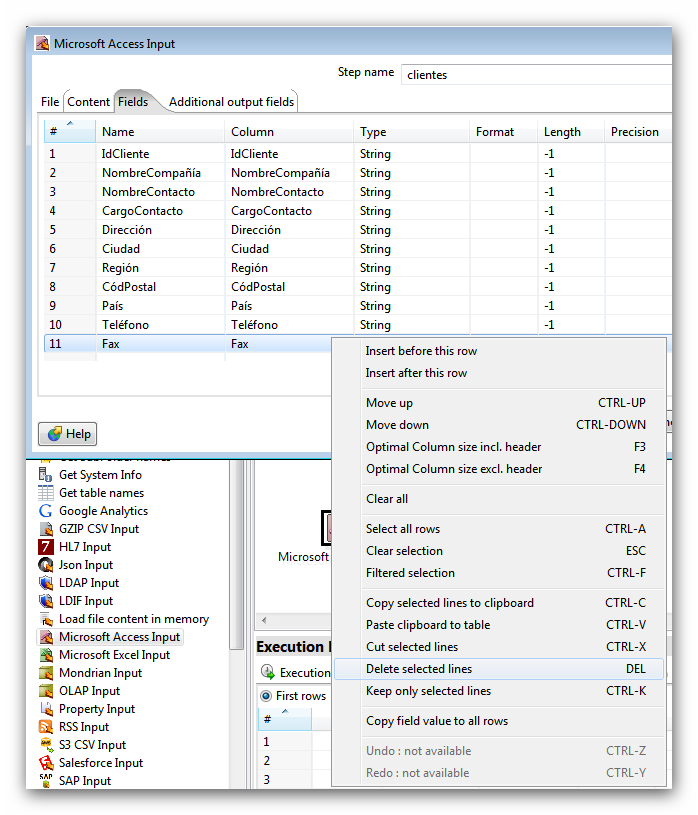
a) Cree un nuevo proceso de transformación y agregue los procesos de tipo **Input -> Microsoft Access Input,** arrastrando el ícono correspondiente al panel principal, y editando sus propiedades para que se pueda conectar a la fuente de datos *Ventas* de la base de datos de Access descargada en el paso 8. Con el botón **Browse** localice la ruta en donde está el archivo *ventas.mdb* (observe la figura siguiente) y posteriormente con el botón **Add** se agrega al panel de archivos elegidos.



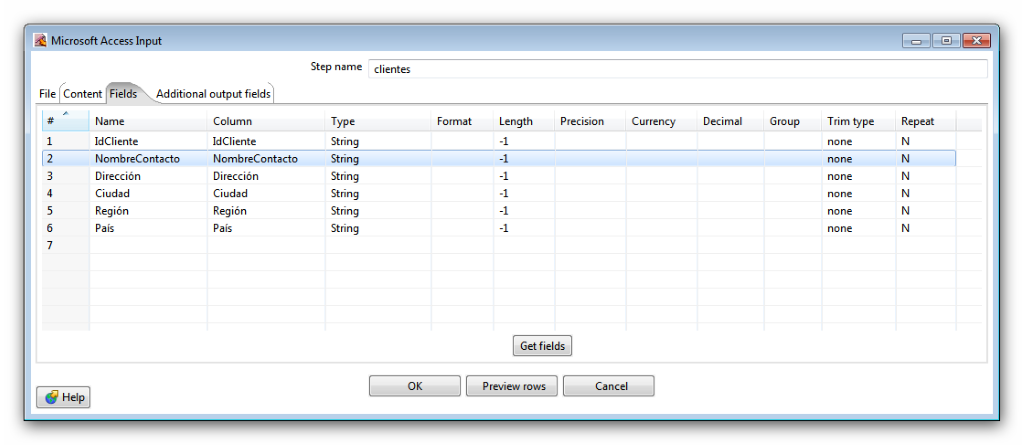
En la pestaña **Content**, oprima el botón **Get Tables**, y elija del cuadro desplegable **Table** la tabla *cliente*.



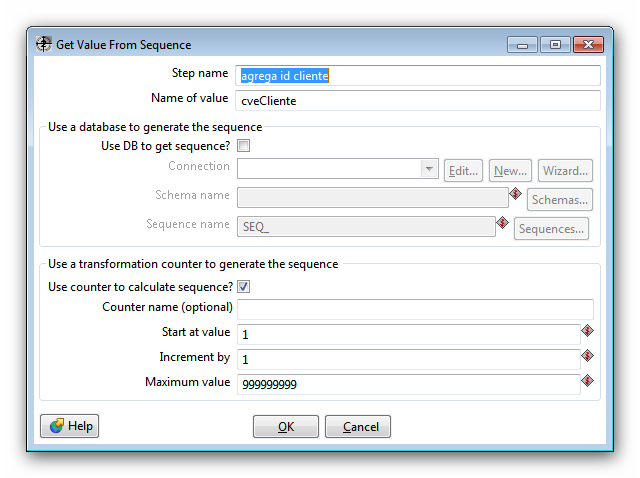
b) En la pestaña **Fields**, oprima el botón **Get Fields**, y observe cómo se llena el panel con los datos de las columnas de la base de datos de Access. De las columnas de la tabla *Cliente*, descarte aquellas que no son necesarias para el modelo multidimensional, con la tecla SUPR (o la opción **Delete selected lines** del menú contextual). Mantenga únicamente las columnas de *IDCliente, NombreCompañia, NombreContacto, CargoContacto, Dirección, Ciudad, Región,* y *País.*



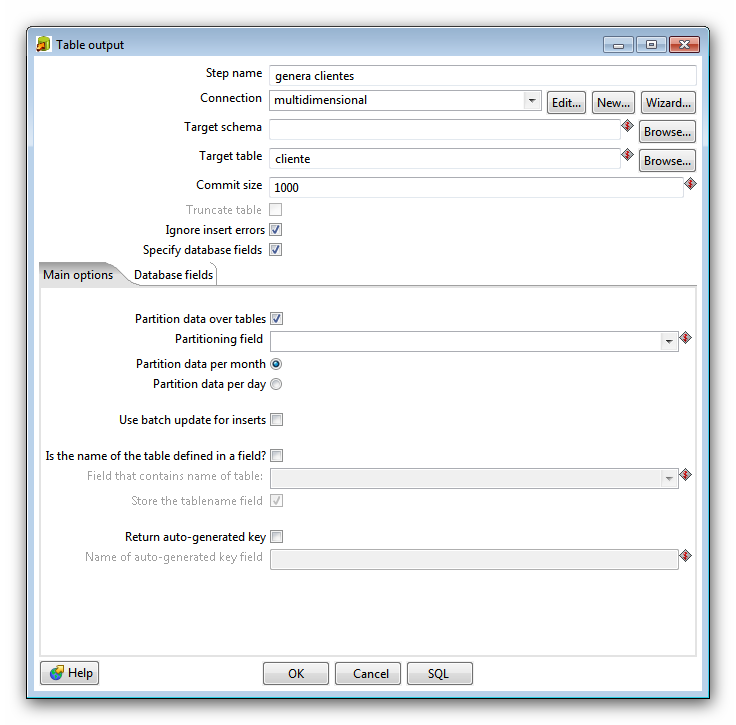
Las columnas de la tabla *Cliente* deben quedar como se muestra en la siguiente figura:



c) Agregue al flujo del proceso un icono de **Add Sequence** y configúrelo (conéctelo con el componente anterior y dé doble clic en él) para que genere una serie de valores secuenciales que serán los identificadores de los clientes. Escriba en el campo de texto **Name of value** el texto *cveCliente*, y verifique que los parámetros coincidan como en la siguiente figura (valores iniciales en 1, incremento en 1 y valor máximo en 9999999999). Dé clic en **OK**.



d) Agregue un icono de **Table output (Salida tabla)** para enviar el proceso como salida a la conexión del servidor Oracle (en la base de datos creada de la práctica anterior), usando el controlador JDBC correspondiente y con los valores de conexión apropiados. Haga el enlace con el componente anterior, y dé doble clic para configurarlo. Oprima el botón **New…** para establecer una conexión a la base de datos de Oracle. (NOTA: Si el checkbox **Partition data over tables (Repartir información en varias tablas)** está seleccionado, dé clic para deseleccionarlo y evitar errores posteriores)



Establezca los valores siguientes en los campos de texto correspondientes:

*Connection name*: multidimensional

*Host name*: localhost

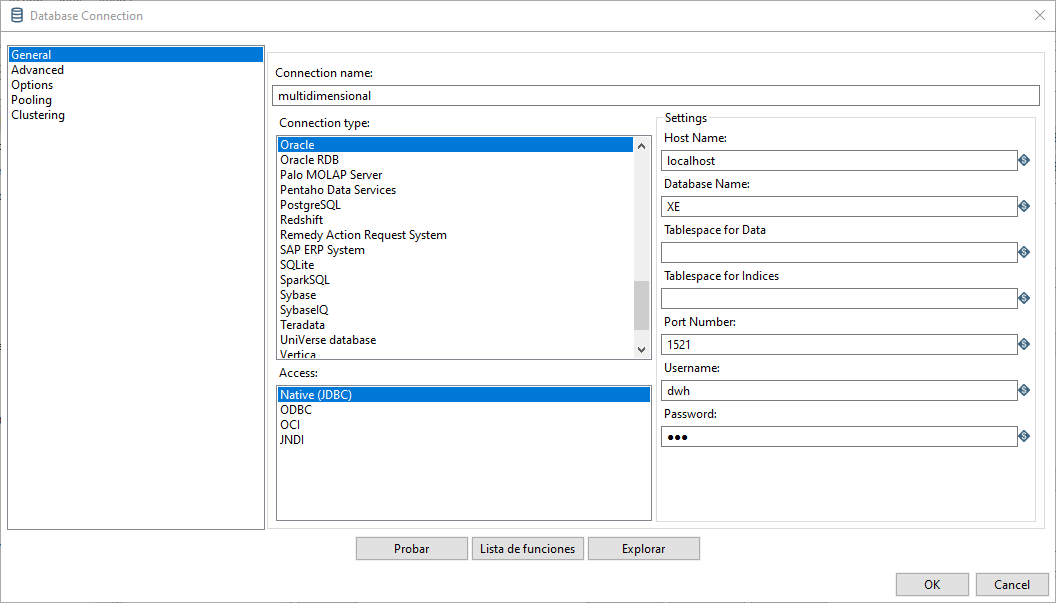
*Database name*: XE

*Port number*: 1521

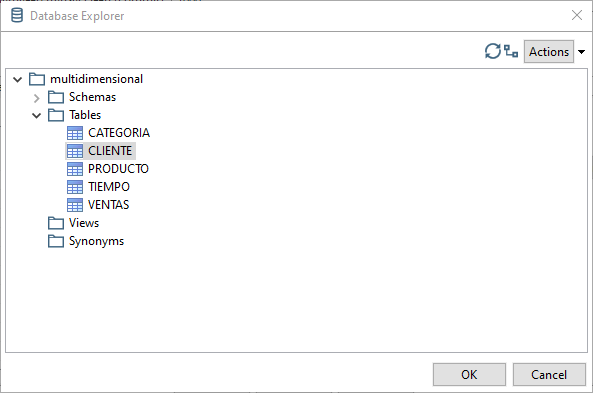
*Username*: dwh

*Password*: dwh

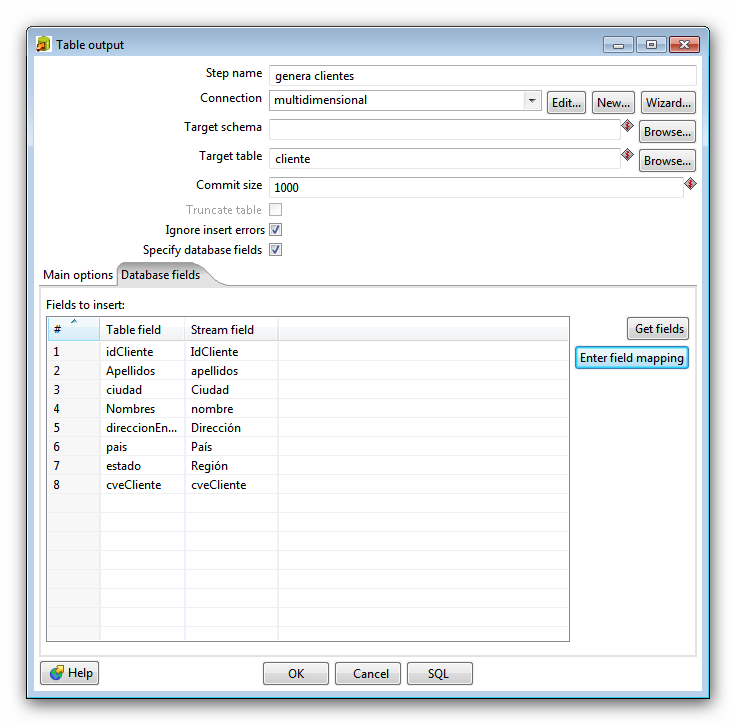
Elija el tipo de conexión como *Oracle* y el acceso como *Native (JDBC)*. Oprima el botón **Probar**, para verificar que se ha establecido una conexión con la base de datos. Si todo es normal, oprima el botón **OK**.



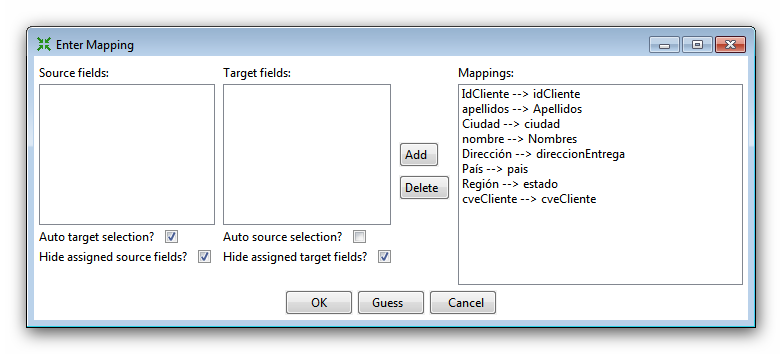
Elija el botón **Examinar…** del campo de texto **Target table (tabla destino)** y de la nueva ventana que aparece, navegue en el árbol del esquema hasta la opción *Tables*, y elija el valor *CLIENTE*, como se muestra en la siguiente figura. De clic en **OK.**



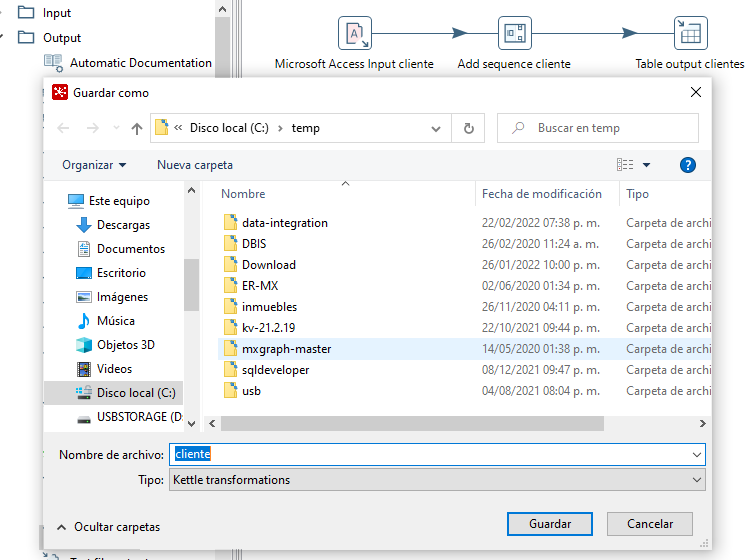
En la pestaña **Database fields**, seleccione el botón **Get Fields**, y observe el conjunto de columnas obtenidas.



Oprima el botón **Enter field mapping**, y de la ventana que aparece, elija cada una de las columnas del panel izquierdo **(Source fields)** con la columna de igual nombre del panel derecho **(Target fields)**, dando clic en el botón **Add** para cada par de columnas. El mapeo debe quedar como en la siguiente figura. Dé clic en **OK**.



e) Guarde el proceso de transformación como un archivo *clientes.ktr* en alguna carpeta de la computadora local, y ejecútelo; observe los resultados.



Incluya una captura de pantalla del proceso creado, incluyendo el *preview* de los datos generados de forma correcta.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

f) En la ventana de comandos de Oracle del punto 4, ejecute la sentencia:

SELECT \* FROM cliente;

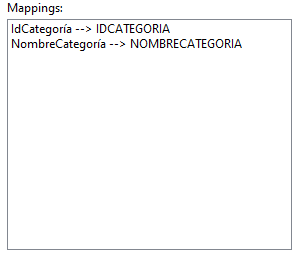
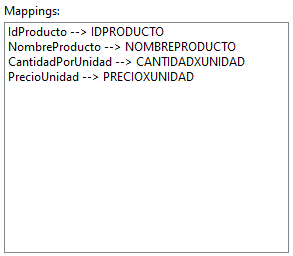
Incluya una captura de los resultados obtenidos.

Texto

Descripción generada automáticamente

Tome en consideración que, si intenta ejecutar nuevamente el proceso anterior, se generará un error en Oracle, debido a que las tablas ya tienen datos y se intenta duplicar valores, por la restricción de llave primaria.

g) Repita el proceso anterior, de manera similar, para las tablas de *producto* y *categoría,* en procesos nuevos y diferentes, considerando los esquemas de las tablas origen (Access) y destino (Oracle). No incluya el componente ***Add Sequence***; únicamente mantenga los componentes **Access** **Input** y **Table Output**. Realice los siguientes mapeos, para las tablas *producto* y *categoría*:



Guarde cada uno de los procesos generados con los nombres correspondientes (*producto.ktr* y *categoria.ktr*), en la misma carpeta de almacenamiento del primer proceso

Incluya una captura de pantalla de los procesos anteriores creados, incluyendo el *preview* de los datos generados de forma correcta. De igual manera, ejecute las sentencias **SELECT \* FROM producto** y **SELECT \* FROM categoria**, de la ventana de comandos de Oracle, e incluya una captura de los resultados mostrados.

Texto

Descripción generada automáticamentePRODUCTOS

Tabla

Descripción generada automáticamente

CATEGORIA

Tabla

Descripción generada automáticamenteTexto

Descripción generada automáticamente

CUESTIONARIO

1.- Describa claramente las características de una bodega de datos (Data Warehouse).

Es específicamente realizado para consultas y análisis.

Es una bodega gigante de datos centralizados

Los datos deben de ser consistentes

Los datos deben de ser organizados de acuerdo a sus temas y resumidos y optimizados

Es variante en el tiempo

Son integradas

2.- Describa las consideraciones a tomar en un esquema multdimensional con respecto a un modelo relacional.

Se debe de tener una tabla de hecho y otras de dimensiones, la información relevante para el negocio se pone en la tabla de hechos.

Las multidimensionales están creadas mediante OLAP, son más versátiles que las relacionales y nos permiten realizar consultas. Generan una búsqueda de parámetros de manera más fácil y sin necesidad de generar muchos joins para poder tener consultas buenas.

Es necesario una tabla de tiempo para generar una tabla que nos ayude con esta restricción

CONCLUSIONES

Escriba sus conclusiones (no personales) con respecto al desarrollo de esta práctica.

Los modelos multidimensionales nos ayudan a tener consultas más sencillas y sin tantos joins. Los modelos multidimensionales tienen mayor eficiencia.

Además, permiten un rendimiento sumamente rápido de las consultas y más si son datos de empresas que contienen datos y más datos, tanto históricos como actuales.

Por otro lado, también permite acceder a grandes cantidades de información y tener el análisis de las relaciones de distintos tipos de elementos relevantes para el negocio

REFERENCIAS BIBLIOGRÁFICAS

1.- Ralph Kimball: “The Data Warehouse Toolkit:

Practical Techniques for Building Dimensional Data Warehouses”

John WIley & Sons, 1996. ISBN 0-471-15337-0

2.- C.J. Date: “Introducción a los sistemas de bases de datos”

Pearson Educación, 2001. ISBN 968-444-419.

3.- Ramez A. Elmasri & Shamkant B. Navathe: “Fundamentos de Sistemas de Bases de Datos”.

Addison-Wesley, 2002 [3ª edición]. ISBN 84-782-9051-6.

4.- Henry F. Korth, Abraham Silberschatz & S. Sudarshan: “Fundamentos de Bases de Datos”.

Mc-Graw Hill, 2002 [4ª edición]. ISBN 84-481-3654-3.